

Applied A.I. Solutions

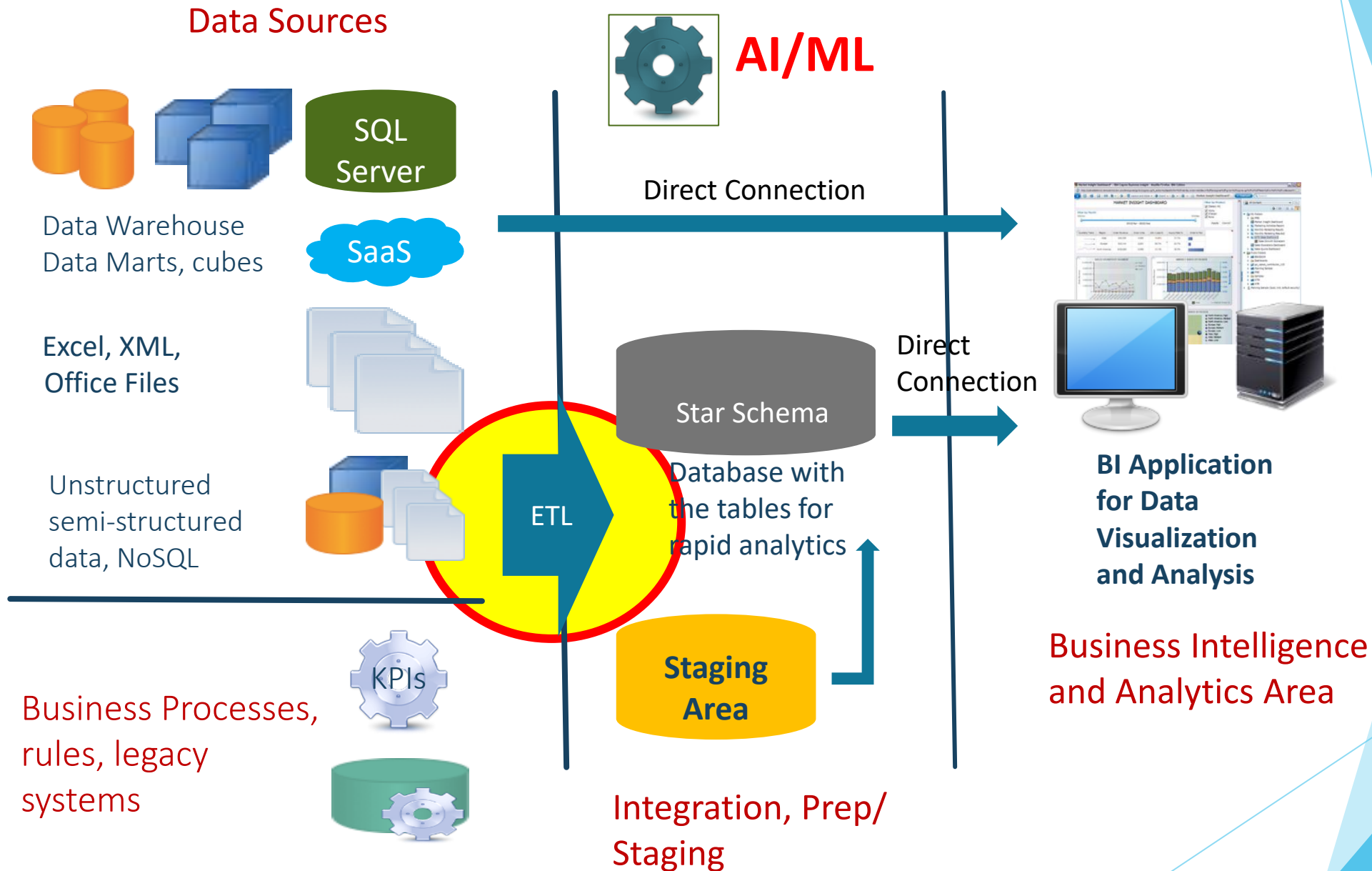
Data Visualization Techniques

Professor
Daniel Vitaver, EMBA

daniel.vitaver@georgebrown.ca

DATA VISUALIZATION TECHNIQUES

Data Preparation (ETL/ELT)



ETL Definition

- An ETL software lets you easily extract data from legacy or modern systems that are not compatible with your infrastructure, transform it into a compatible format, and load it to the destination system.

ETL lifecycle

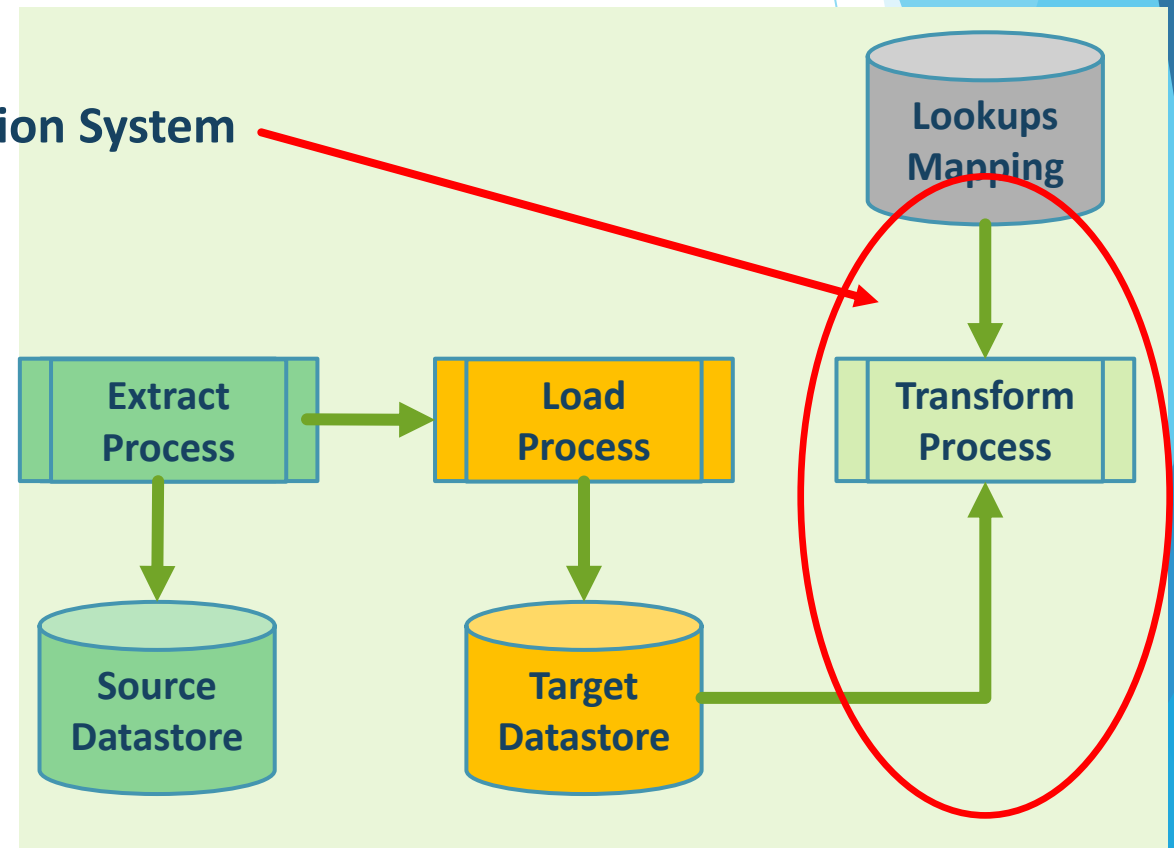
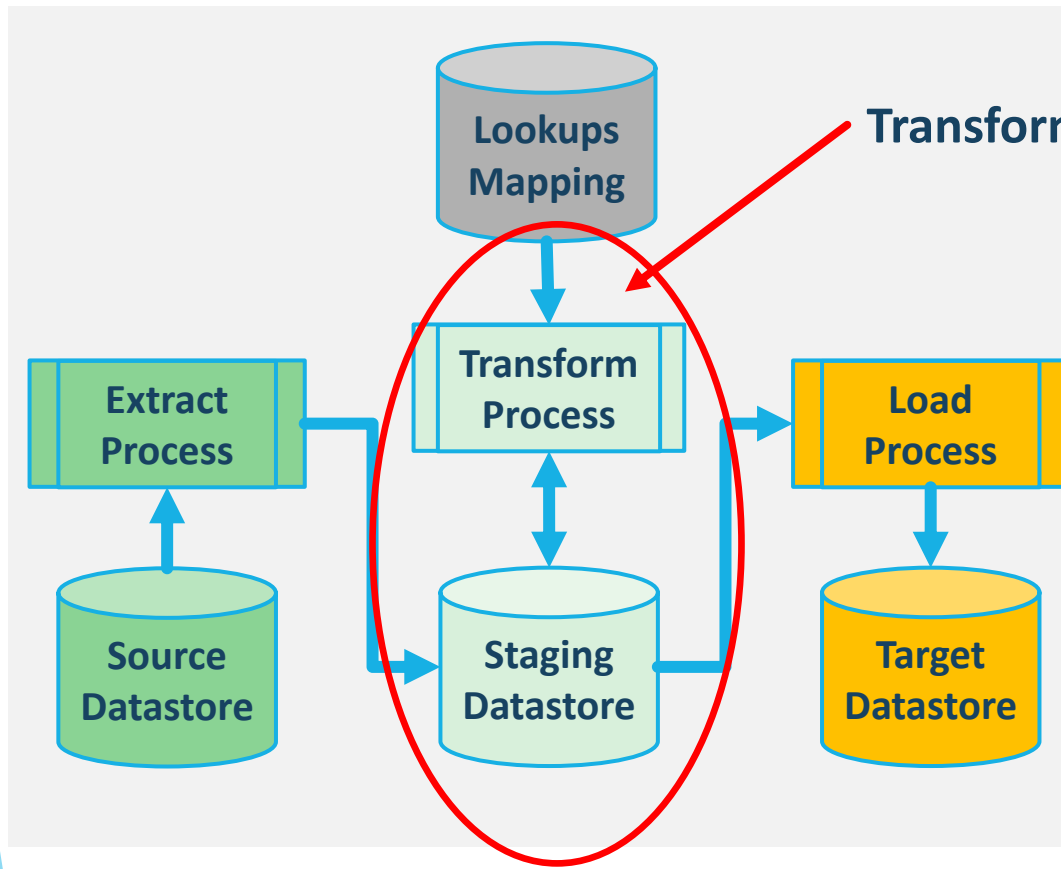
1. Cycle initiation
2. Build reference data
3. Extract (from sources)
4. Validate
5. Transform (clean, apply business rules, check for data integrity, create aggregates or disaggregates)
6. Stage (load into staging tables, if used)
7. Audit reports (for example, on compliance with business rules. Also, in case of failure, helps to diagnose/repair)
8. Publish (to target tables)
9. Archive

- **ETL, Data Prep, Data Enhancement - Main features**
 - Based on a defined **set of rules**. No-code ETL tools
 - Data Discovery, and Data Cleansing
 - Data Mapping, Data Lineage Flows and Data Transformation
 - Data Integration
 - Data Synchronization
 - Workflow Job Scheduling and Automation
 - Enforce policies set by a Data Governance Program

Data Cleansing Definition

- Data cleansing or data cleaning is the **process of detecting and correcting** (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

ETL vs ELT



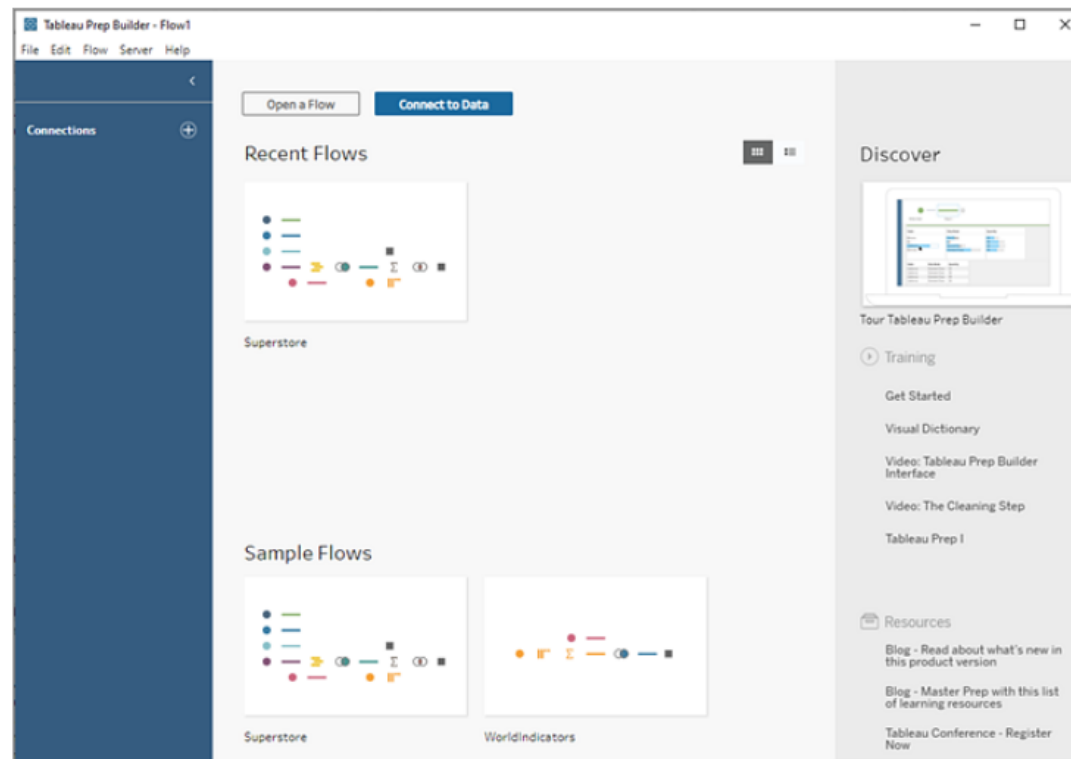
DATA VISUALIZATION TECHNIQUES

Tableau Software

Tableau Prep Builder

1. Connect to data

The first thing you see when you open Tableau Prep Builder is a Start page with a **Connections** pane, just like Tableau Desktop.

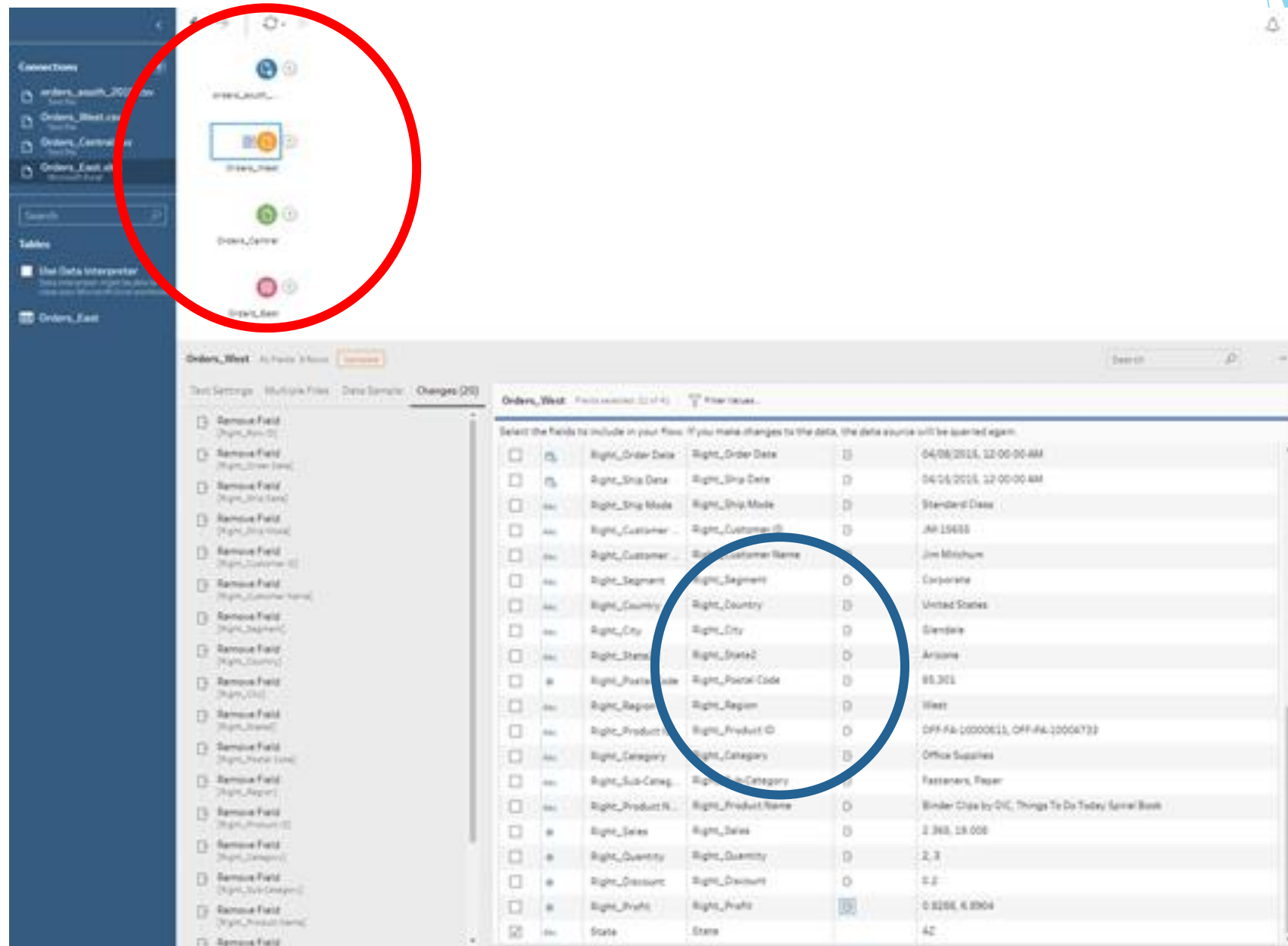


— Input Pane

The screenshot shows the Microsoft Power BI Desktop interface. On the left, the 'Home' ribbon is visible, and the 'Data' source is set to 'orders_south_2015.csv'. The main area displays the 'Import' dialog box, which is open to the 'Multiple Files' tab. A red circle highlights the 'Import' button. The dialog shows a list of files to be imported, including 'orders_south_2015.csv', 'orders_south_2016.csv', 'orders_south_2017.csv', and 'orders_south_2018.csv'. The 'Import' button is highlighted with a red circle.

2. Explore your data

- Now that you have the data files loaded into Tableau Prep, you're pretty sure that you want to combine all of the files together. But before you do that, it might be a good idea to **take a look at them first and see if you can spot any issues.**
- When you select an Input step in the Flow pane, you can see the settings used to bring in the data, the fields that are included, and a preview of your values.
- This is a good place to decide **how much data** you want to include in your flow and **remove or filter fields** that you don't want. You can also **change any data types** that were assigned incorrectly.



The screenshot displays the Microsoft Power BI Desktop interface. On the left, the 'Connections' pane shows a list of data sources, with 'Orders_West' highlighted by a red circle. Below this, the 'Fields' pane shows a list of fields for the 'Orders_West' data source, with 'Orders_West' highlighted by a blue circle. The main area of the screen shows a table of data for 'Orders_West'.

Field	Value
Right_Order Date	04/08/2015, 12:00:00 AM
Right_Ship Date	04/08/2015, 12:00:00 AM
Right_Ship Mode	Standard Class
Right_Customer ID	NP 15655
Right_Customer Name	Jim Michum
Right_Segment	Corporate
Right_Country	United States
Right_City	Glendale
Right_State	Arizona
Right_Postal Code	85301
Right_Region	West
Right_Product ID	OFF-PA-10000611, OFF-PA-10006723
Right_Category	Office Supplies
Right_Sub-Category	Pasteners, Paper
Right_Product Name	Binder Clips by DIX, Things To Do Today Spiral Book
Right_Sales	2,363,18,000
Right_Quantity	2,3
Right_Discount	0.2
Right_Profit	0.0206, 6.8904
State	42

3. Clean your data

- In Tableau Prep, examining and cleaning your data is an iterative process.
- After you decide on the data set that you want to work with, the next step is to examine and take action on that data by applying various cleaning, shaping, and combining operations to it.
- You apply these operations by adding steps to your flow.

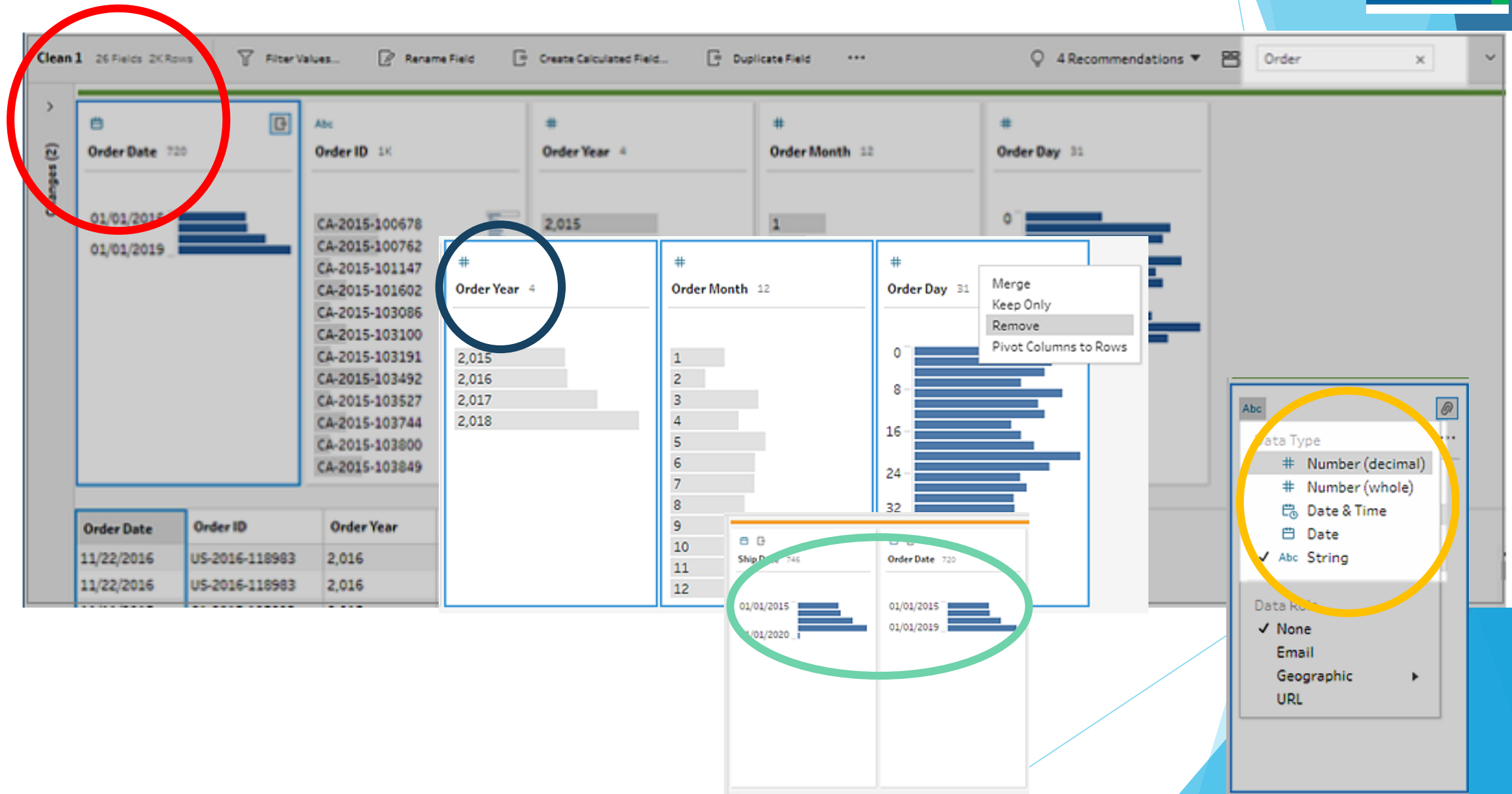


Tableau Dashboard Overview:

- Order Date (720):** Bar chart showing counts for 01/01/2015 and 01/01/2019. (Circled in red)
- Order ID (1K):** Bar chart showing counts for various IDs including CA-2015-100678, CA-2015-100762, CA-2015-101147, CA-2015-101602, CA-2015-103086, CA-2015-103100, CA-2015-103191, CA-2015-103492, CA-2015-103527, CA-2015-103744, CA-2015-103800, and CA-2015-103849.
- Order Year (4):** Bar chart showing counts for 2,015, 2,016, 2,017, and 2,018. (Circled in blue)
- Order Month (12):** Bar chart showing counts for months 1 through 12.
- Order Day (31):** Bar chart showing counts for days 0, 8, 16, 24, and 32.

Context Menu (Order Day):

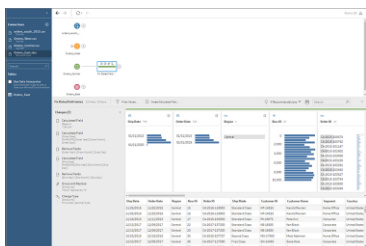
- Merge
- Keep Only
- Remove
- Pivot Columns to Rows

Data Type Dropdown (Order Date):

- Number (decimal)
- Number (whole)
- Date & Time
- Date
- ☒ String
- None
- Email
- Geographic
- URL

Tableau Interface Elements:

- Top Bar: Clean 1, 26 Fields, 2K Rows, Filter Values..., Rename Field, Create Calculated Field..., Duplicate Field, 4 Recommendations, Order
- Left Panel: Changes (2)
- Bottom Panel: Order Date, Order ID, Order Year



View and clean data



100%

Clean 1 24 fields 2K rows

Filter Values...

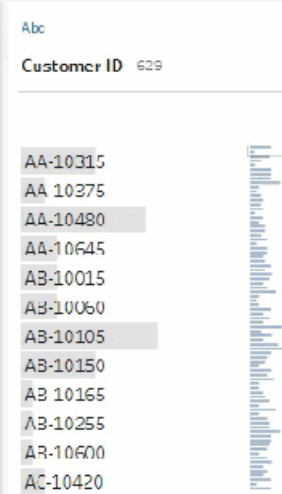
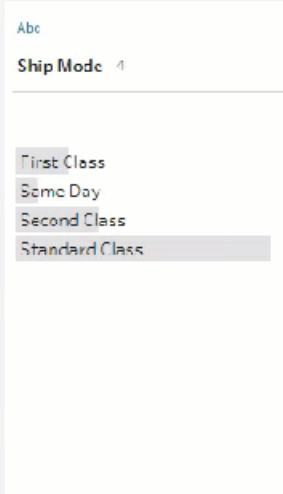
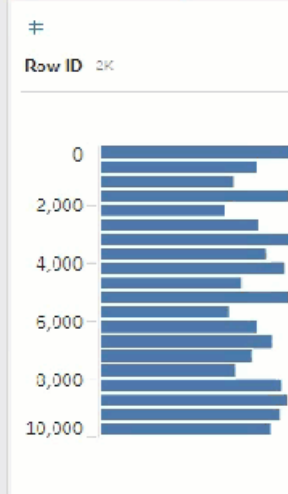
Create Calculated Field...

4 Recommendations

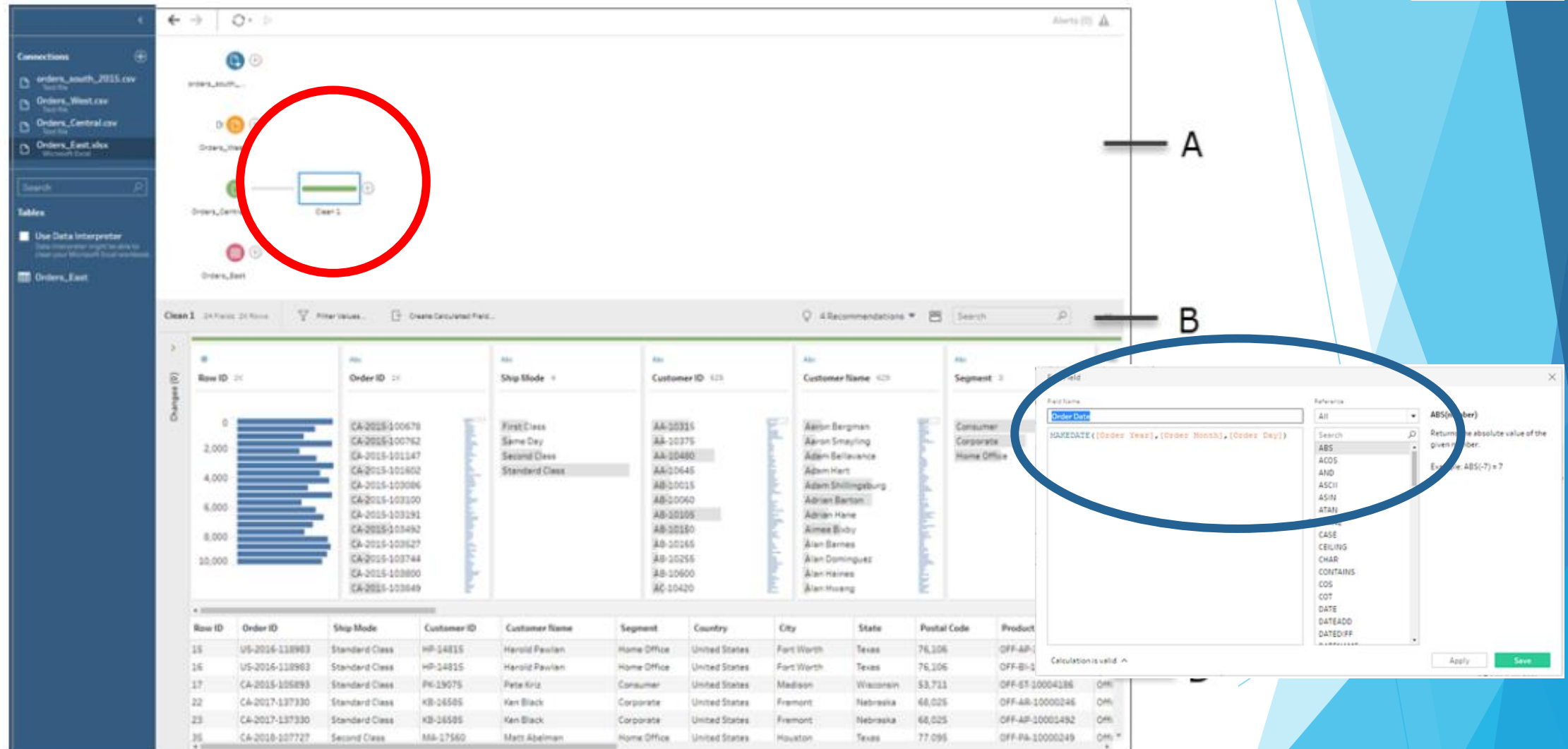
Search

Changes (0)

No data changes.
Clean and transform your data and see a list of your changes here.



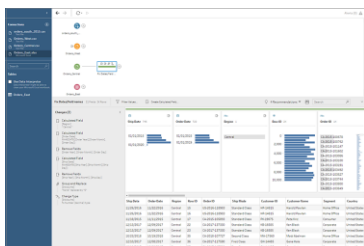
Row ID	Order ID	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Product ID	Category
15	US-2016-118983	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76,106	OFF-AP-10002311	Office Supply
16	US-2016-118983	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76,106	OFF-BI-10000756	Office Supply



The screenshot displays the Power BI Desktop interface. On the left, the 'Connections' pane shows several CSV files. The main area is divided into four panes: the 'Flow pane' (top left, circled in red), the 'Toolbar' (top right, labeled 'A'), the 'Profile pane' (bottom left, labeled 'B'), and the 'Data grid' (bottom right). The 'Flow pane' shows a diagram of data sources and a calculated field named 'Clean 1'. The 'Toolbar' contains various icons for data manipulation. The 'Profile pane' shows a bar chart and a table of data. The 'Data grid' shows a table of data with columns for Row ID, Order ID, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, and Product.

Label A points to the Toolbar. Label B points to the Profile pane.

A. Flow pane, B. Toolbar, C. Profile pane, D. Data grid



View and clean data



100%

Clean 1 24 fields 2K rows

Filter Values...

Create Calculated Field...

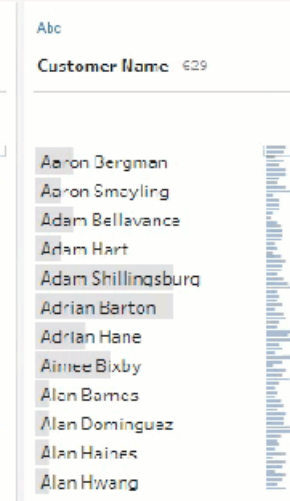
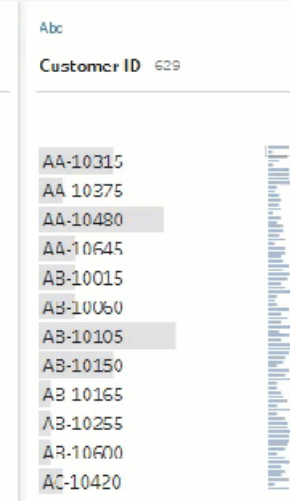
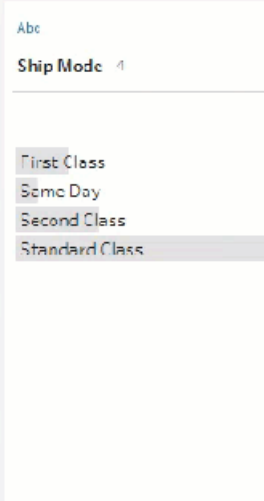
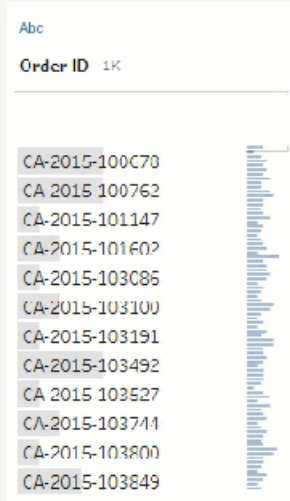
4 Recommendations

Search

Changes (0)

No data changes.

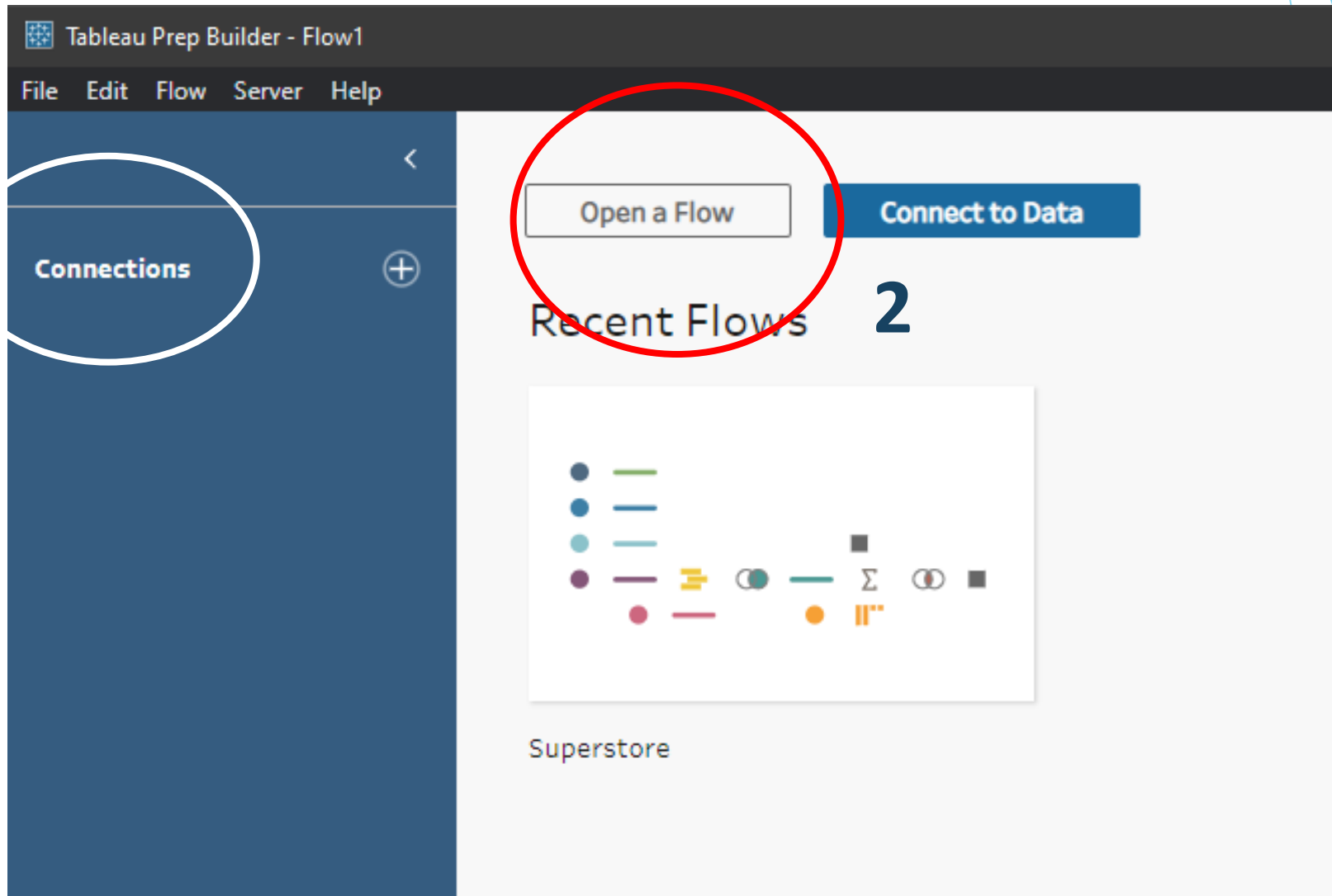
Clean and transform your data and see a list of your changes here.



Row ID	Order ID	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Product ID	Category
15	US-2016-118983	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76,106	OFF-AP-10002311	Office Supply
16	US-2016-118983	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76,106	OFF-BI-10000756	Office Supply

TABLEAU – Data Preparation Builder

1



2

TABLEAU – Data Flow

1

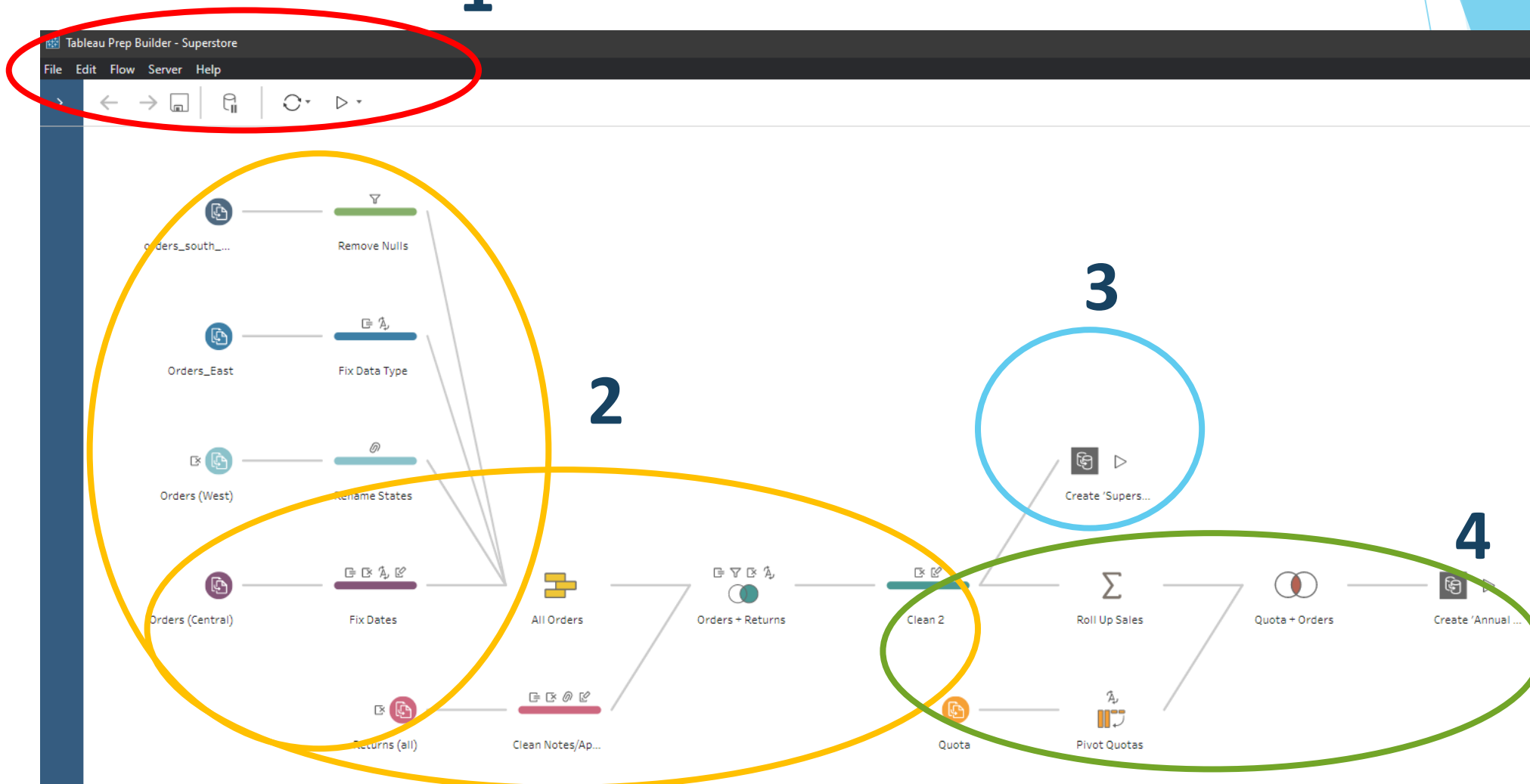


TABLEAU – Data Flow – Data Transformation

1

2

3

4

Connections

- Orders_Central.csv (Text file)
- Orders_West.csv (Text file)
- return reasons_new... (Microsoft Excel)
- Quota.xlsx (Microsoft Excel)
- Orders_East.xlsx (Microsoft Excel)

Data Flow Diagram:

Orders (Central) → Fix Dates → All Orders → Orders + Returns → Clean 2 → Roll Up Sales → Quota + Orders → Create 'Annual ...'

Orders + Returns (29 fields, 16K rows)

Settings

- Change Type: [Discount] To Number (decimal) type
- Calculated Field: [Days to Ship] DATEDIFF('day', [Order Date], [Ship Date])
- Calculated Field: [Returned?] IF ISNULL([Return Reason]) = false THEN "Yes" ELSE "No" END
- Remove Field: [Table Names]
- Remove Field: [File Paths]
- Calculated Field: [Discount] IFNULL([Discount], 0)
- Calculated Field: [Year of Sale] YEAR([Order Date])
- Filter: [Discount] Exclude: (17.0 - 18.0)
- Join: [Product ID] == [Product ID], [Order ID] == [Order ID]

Join Clauses

Clean Notes/Approver

Product ID	Order ID
FUR-BO-10000362	CA-2015
FUR-BO-10002268	CA-2015
FUR-BO-10003150	CA-2016
FUR-BO-10004218	CA-2015
FUR-CH-10000847	CA-2017
FUR-CH-10000847	CA-2017
FUR-CH-10000863	CA-2018
FUR-CH-10000908	CA-2015
FUR-CH-10001146	CA-2015
FUR-CH-10002647	CA-2016
FUR-CH-10002774	CA-2018
FUR-CH-10002961	CA-2018
FUR-CH-10003396	CA-2016
FUR-CH-10003535	CA-2016
FUR-CH-10003761	CA-2015
FUR-CH-10003833	US-2017

All Orders

Product ID	Order ID
FUR-BO-10000112	CA-2018
FUR-BO-10000330	CA-2015
FUR-BO-10000330	CA-2016
FUR-BO-10000330	CA-2018
FUR-BO-10000362	CA-2015
FUR-BO-10000362	CA-2015
FUR-BO-10000362	CA-2016
FUR-BO-10000362	CA-2017
FUR-BO-10000362	US-2015
FUR-BO-10000468	CA-2015
FUR-BO-10000468	CA-2016
FUR-BO-10000468	CA-2018
FUR-BO-10000468	CA-2018
FUR-BO-10000468	CA-2018
FUR-BO-10000711	CA-2015

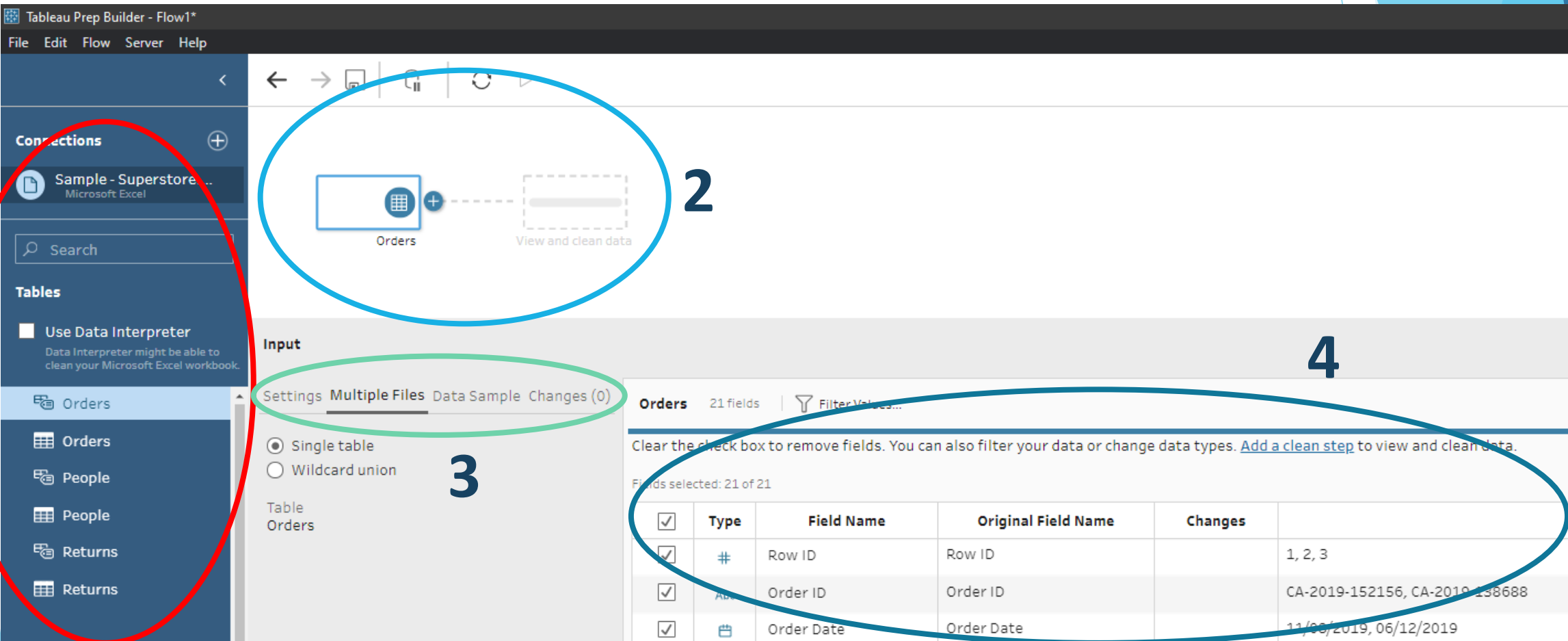
Join Results

Year of Sale	Returned?	Days to Ship
2,015	null	0
2,016	Yes	1
2,017		2
2,018		3
		4
		5
		6
		7

Table View:

Year of Sale	Returned?	Days to Ship	Approver	Return Notes	Order ID	Product ID
2,016	null	4	null	null	null	null
2,016	null	4	null	null	null	null

TABLEAU – Data Cleansing



The screenshot shows the Tableau Prep Builder interface. The left sidebar (1) contains the 'Connections' pane with 'Sample - Superstore...' and the 'Tables' pane with 'Orders', 'People', 'Returns', and 'Returns'. The main canvas (2) shows a flow starting with an 'Orders' data source icon followed by a 'View and clean data' step icon. The bottom section (3) shows the 'Input' tab with 'Settings', 'Multiple Files', 'Data Sample', and 'Changes (0)' sub-tabs. The 'Multiple Files' sub-tab is selected, showing 'Single table' and 'Wildcard union' options. The right pane (4) shows the 'Orders' data source with 21 fields, a 'Filter Values...' button, and a table of field details.

Field Name	Original Field Name	Changes
Row ID	Row ID	1, 2, 3
Order ID	Order ID	CA-2019-152156, CA-2019-158688
Order Date	Order Date	11/08/2019, 06/12/2019

1

2

3

4

Steps

1. Connect to data
2. Explore your data
3. Clean your data
4. Combine your data
 - Union data
 - Clean data
 - Join your data
 - Clean join results
 - Name step
 - Save flow
5. Run flow and generate output

[For details, get started visit Tableau](#)

